Journal of Clinical Densitometry: Assessment & Management of Musculoskeletal Health, vol. 19, no. 3, 332–339, 2016

1094-6950/19:332–339/$36.00
http://dx.doi.org/10.1016/j.jocd.2016.02.001

Original Article

Evaluation of Automated Fracture Risk Assessment Based on the Canadian Association of Radiologists and Osteoporosis Canada Assessment Tool

Sonya Allin,[*,1] *Robert Bleakney,*[2,3,4] *Julie Zhang,*[1] *Sarah Munce,*[1] *Angela M. Cheung,*[2,3,4,5] *and Susan Jaglal*[1,6]

[1]*Department of Physical Therapy, University of Toronto, Toronto, ON, Canada;* [2]*Department of Medical Imaging, Mount Sinai Hospital and the University Health Network, Toronto, ON, Canada;* [3]*Joint Department of Medical Imaging (JDMI), University Health Network, Toronto, ON, Canada;* [4]*Centre of Excellence in Skeletal Health Assessment, University of Toronto, Toronto, ON, Canada;* [5]*University Health Network Osteoporosis Program, Toronto General Hospital, Toronto, ON Canada; and* [6]*Osteoporosis Research Program, Women's College Research Institute, Toronto, ON, Canada*

Abstract

Fracture risk assessments are not always clearly communicated on bone mineral density (BMD) reports; evidence suggests that structured reporting (SR) tools may improve report clarity. The aim of this study is to compare fracture risk assessments automatically assigned by SR software in accordance with Canadian Association of Radiologists and Osteoporosis Canada (CAROC) recommendations to assessments from experts on narrative BMD reports. Charts for 500 adult patients who recently received a BMD exam were sampled from across University of Toronto's Joint Department of Medical Imaging. BMD measures and clinical details were manually abstracted from charts and were used to create structured reports with assessments generated by a software implementation of CAROC recommendations. CAROC calculations were statistically compared to experts' original assessments using percentage agreement (PA) and Krippendorff's alpha. Canadian FRAX calculations were also compared to experts', where possible. A total of 25 (5.0%) reported assessments did not conform to categorizations recommended by Canadian guidelines. Across the remainder, the Krippendorff's alpha relating software assigned assessments to physicians was high at 0.918; PA was 94.3%. Lower agreement was associated with reports for patients with documented modifying factors (alpha = 0.860, PA = 90.2%). Similar patterns of agreement related expert assessments to FRAX calculations, although statistics of agreement were lower. Categories of disagreement were defined by (1) gray areas in current guidelines, (2) margins of assessment categorizations, (3) dictation/transcription errors, (4) patients on low doses of steroids, and (5) ambiguous documentation of modifying factors. Results suggest that SR software can produce fracture risk assessments that agree with experts on most routine, adult BMD exams. Results also highlight situations where experts tend to diverge from guidelines and illustrate the potential for SR software to (1) reduce variability in, (2) ameliorate errors in, and (3) improve clarity of routine adult BMD exam reports.

Key Words: Dual-energy X-ray; guidelines; osteoporosis; radiology information systems; structured reporting.

Introduction

A bone mineral density (BMD) exam measures the quantity of minerals in bone and provides an assessment of 10-yr fracture risk. Both the Canadian Association of Radiologists (CAR) and the International Society for Clinical Densitometry suggest that fracture risk assessments

Received 10/19/15; Revised 02/9/16; Accepted 02/16/16.
*Address correspondence to: Sonya Allin, PhD, Department of Physical Therapy, University of Toronto, 160–500 University Avenue, Toronto, ON M5G 1V7, Canada. E-mail: s.allin@utoronto.ca

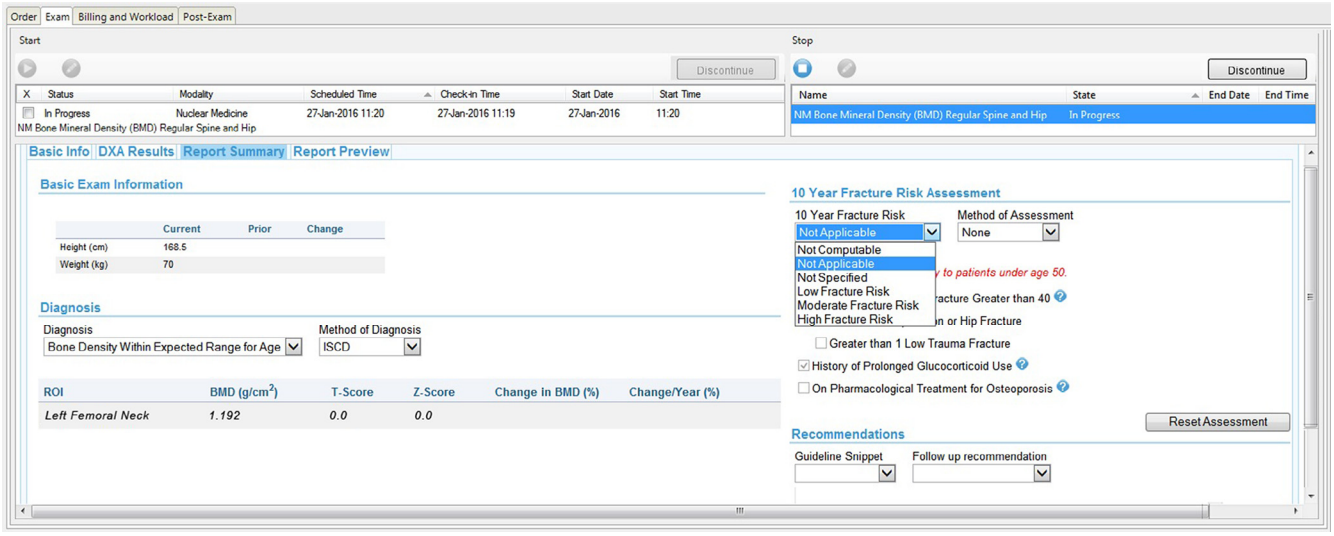

Fig. 1. A screenshot illustrating automated suggestions for risk and diagnosis from the Joint Department of Medical Imaging Radiological Information System's BMD reporting software. Technologists input clinical details, and software makes suggestions for assessment and diagnosis based on these details. Software suggestions can be overridden by reading physicians, where necessary. BMD, bone mineral density; ROI, region of interest.

appear on all BMD reports *(1–5)*; many osteoporosis guidelines additionally suggest that they be used to guide clinicians' treatment decisions *(5,6)*. A popular risk assessment tool is the internationally validated FRAX, which estimates fracture risk based on measured BMD, age, sex, history of fragility fracture, history of steroid use, and other clinical factors *(7)*. However, the FRAX requires access to proprietary software and/or internet access to use. In Canada, a simplified and accessible alternative assessment tool is called the Canadian Association of Radiologists and Osteoporosis Canada (CAROC) *(8,9)*. Similar to the FRAX, the CAROC attends to age, sex, and BMD results, but does not consider factors like family history *(8)*. Despite its relative simplicity, the CAROC produces categorical assessments of risk (e.g., "high," "moderate," or "low") that agree with the FRAX approximately 89% of the time *(8)*.

When fracture risks are accurate and communicated clearly, they are desirable and influential to referring physicians *(10–12)*. Moreover, when directly related to treatment directives, they increase the likelihood that appropriate treatment decisions will follow *(11–13)*. However, evidence *also* indicates that fracture risk assessments on BMD reports are often communicated poorly *(14,15)*. Research suggests family physicians to be routinely uncertain as to clinical details or methods informing fracture risks on reports *(14)*; surveys of reports have demonstrated lack of detail as to the "ingredients" of risk calculations *(16)*. It is, however, important for referring physicians to have insight into assessment calculations as missing ingredients may influence risk accuracy. This is made evident by recent studies of reports for known fracture patients, wherein more than half omitted fracture history and, consequently, underestimated reported risk *(16,17)*. Complicating the situation further is the fact that BMD reports often appear as short narratives; reports with this structure have been shown to be more confusing to referring physicians than those that are longer in length *(13)*.

In an effort to improve the quality of its BMD reports, the Joint Department of Medical Imaging (JDMI) of Toronto's University Health Network, Women's College Hospital, and Mount Sinai Hospital is experimenting with a custom structured reporting (SR) solution for BMD that is integrated with their Radiological Information System (RIS). This solution automatically suggests CAROC-consistent fracture risk assessments based on manual entry of patient details. Final structured reports detail assessment algorithms and clinical factors used in risk calculation by default. JDMI's move to this SR system is grounded in substantial evidence that demonstrates SR to enhance the clarity and completeness of diagnostic imaging reports *(18,19)*. The additional inclusion of CAROC calculations is supported by research validating the accuracy of automated assessment calculations in other diagnostic imaging domains *(20,21)* and their benefit by reading clinicians, particularly those with limited domain familiarity *(22)*.

A screenshot illustrating automated suggestions for risk as they take place within the RIS is presented in Fig. 1. Figure 2 provides a pseudocode for the algorithm that is used to suggest risk, and Fig. 3 provides the system's structured format for a completed report.

The objectives of the present study are (1) to determine the ability of JDMI's SR system to assign fracture risk assessments that agree with JDMI experts and (2) to identify clinical situations that prompt divergence from auto-

Input: BMD T-scores from lumbar spine, left femoral neck, right femoral neck, left total hip and right total hip (where these exist), age, gender, risk factors as T/F variables (history of fragility fracture after 40 and history of steroid use for 3 months or more in past year at 7.5mg Prednisone equivalent daily)

Output: Categorical assessment of 10-year Fracture Risk according to CAROC 2010 guidelines (High, Moderate, Low, Undefined or Not Applicable)

If (Patient > 85): Assess patient as though they were of age 85.

If (Patient < 50): Set Fracture Risk to Not Applicable.

If (Patient >= 50):

If Patient has a history of Vertebral Fracture or Multiple Fragility Fractures, set Fracture Risk to High.

Otherwise:

1. Set Fracture Risk to Undefined.
2. Select T-score from left femoral neck. (If, however, this T-score is unavailable, substitute T-score from left total hip, right femoral neck, or right total hip.)
3. Given a valid T-score:
   a) Set Fracture risk to Basal CAROC Risk (based on gender, age and T-score).
   b) If history of fragility fracture is TRUE, elevate risk one level (High is maximum).
   c) If history of steroid use is TRUE, elevate risk one level (High is maximum).
4. If Fracture Risk is Low or Undefined and either lumbar spine or total hip t-score exists and is <= -2.5, set Fracture Risk to Moderate.

Output Fracture Risk

Fig. 2. Pseudocode Joint Department of Medical Imaging implementation of CAROC algorithm for 10-yr absolute fracture risk assessment. BMD, bone mineral density; CAROC, Canadian Association of Radiologists and Osteoporosis Canada.

mated risk calculations so that reports can be flagged for close review when such situations occur.

Materials and Methods

Charts for 500 adult patients who received routine BMD exams between January and June 2014 were selected from 4 hospital partners of JDMI, who are members of the University of Toronto's Centre for Excellence in Skeletal Health Assessment. Centre for Excellence in Skeletal Health Assessment physicians responsible for reading BMD exams include osteoporosis specialists and radiologists. A total of 219 reports were sampled for patients of specialists and 281 for patients of radiologists. Charts containing verified routine adult BMD examination results were included; charts containing BMD exam results that were either not finalized by the reporting clinician or conducted for research purposes or for patients under the age of 18 were excluded.

Clinical information of relevance to CAROC assessments were manually abstracted from both free text reports and accompanying patient charts. This information included demographics (e.g., age and sex), BMD results (i.e., raw BMD in gram per square centimeter and BMD *T*-scores and/or *Z*-scores, where applicable), as well as history of fragility fracture and glucocorticoid use. Additional details of relevance to FRAX calculations were also abstracted; these included height, weight, smoking status, history of parental fracture, units of daily alcohol intake, and information on secondary conditions. Experts' fracture risk assessments were recorded in whatever language they appeared on reports. Of note, several BMD reports contained 2 risk assessments: the first being a basal risk based on *T*-score data and the second reflecting additional clinical factors. In these situations, the second assessment was abstracted and modifying factors were assumed. Abstractions were performed by 2 members of the research team using a standardized protocol; data were periodically validated against patient records by the lead author.

BONE MINERAL DENSITY REPORT

DOB: XX/XX/XXXX

Gender: XXX
Current Height: XXX cm
Current Weight: XX kg

Date of Prior Exam: XX/XX/XXXX

*** ABSOLUTE TEN YEAR FRACTURE RISK[1] IS MODERATE (10 TO 20% OVER 10 YEARS) ***

THIS ASSESSMENT IS BASED ON THE 2010 CAROC ASSESSMENT TOOL.

RISK FACTORS INCLUDED:
-- Hx of Low Trauma Fracture: XX
  -- Vertebral Compression or Hip Fracture: XX
  -- More Than 1 Low Trauma Fracture: XX
-- Hx of Steroid Use[2]: XX
-- On Pharmacological Treatment for Osteoporosis: XX

DXA RESULTS	RAW BMD (g/cm2)	T-SCORE	CHANGE SINCE XXXX
LEFT FEMORAL NECK	X.XXX	X.X	X.X%*
TOTAL LEFT HIP	X.XXX	X.X	X.X%
LUMBAR SPINE (L1 - L4)	X.XXX	X.X	X.X%

** indicates statistically significant change*

DIAGNOSIS: Reduced Bone Mineral Density[3]

[Any dictated notes from the reading physician appear here; content is left to the physician's discretion.]

Factors that Warrant considerations for Pharmacological Therapy according to the 2010 Osteoporosis Canada Guidelines:
- Additional vertebral fracture(s) identified;
- A lumbar spine T-score that is much lower than the femoral neck T-score;
- Rapid bone loss as evidenced by prior BMD exams;
- Recurrent falls as defined as falling 2 or more times in the past 12 months;
- Other disorders strongly associated with osteoporosis, bone loss or fracture

1 Fracture risk reported above is for an untreated patient. Patients on pharmacotherapy for osteoporosis will have a lower fracture risk than this reported value. History of glucocorticoid use and low trauma fracture elevate fracture risk. If a known risk factor has been omitted, risk may be under-estimated.
2 >= 3 months in the prior year at a prednisone equivalent dose >= 7.5 mg daily
3 WHO criteria and recommended ISCD terminology used for diagnosis.

Fig. 3. An example of the Radiological Information System's structured BMD report template for a "moderate risk" BMD report; "XXX" indicates a location where patient information can be located in the format. Note that this template has yet to be finalized by the Joint Department of Medical Imaging. BMD, bone mineral density; DOB, date of birth; DXA, dual-energy X-ray absorptiometry.

To calculate CAROC-consistent assessments, the CAROC implementation in JDMI's SR software processed abstracted data offline. To do this, abstracted data were organized into comma-delimited files and batch processed by the SR system's assessment calculation algorithm, presented in Fig. 2. Of note, the implemented algorithm sometimes elevates calculated risk based on data from sites other than the femoral neck *(6)*. Elevations follow an interpretation of Canadian guidelines in the CAR Technical Standards; this elevates risk based on lumbar spine

Table 1
Characteristics of Sampled Reports

Characteristics	Radiologists	Specialists	*p* Value
Reports sampled	281	219	
Average age of patients (standard deviation)	61.0 (12.0)	59.1 (18.2)	
Female, n (%)	225 (80.4)	167 (76.3)	0.02
On bone sparing treatments, n (%)	24 (8.6)	39 (17.8)	0.005
With history of low-trauma fracture, n (%)	26 (9.3)	31 (14.2)	0.34
With >1 low-trauma fractures, n (%)	1 (0.4)	12 (5.5)	0.008
With vertebral fractures, n (%)	0 (0.0)	6 (2.7)	<0.001
On steroids (over 7.5 mg daily), n (%)	10 (3.6)	25 (11.4)	<0.001
On steroids (between 5.0 and 7.5 mg daily), n (%)	11 (3.9)	52 (23.7)	<0.001
Baseline exams, n (%)	81 (40.7)	146 (45.6)	
Experts' fracture risk categorizations, n (%)			
High	13 (4.6)	31 (14.2)	<0.001
Moderate to high	0 (0.0)	10 (4.6)	<0.001
At least moderate	7 (2.1)	0 (0.0)	<0.001
Moderate	60 (21.4)	66 (30.1)	0.18
Low to Moderate	0 (0.0)	8 (3.7)	<0.001
Low	157 (56.1)	47 (21.5)	<0.001
Not applicable	35 (12.5)	57 (26.0)	0.05

or total hip *T*-scores only when patients are "low risk" according to femoral neck data *(3)*.

Where possible, abstracted data were also manually provided to the freely available online Canadian FRAX calculator *(23)*. However, the FRAX requires knowledge of some clinical variables that are not required by the CAROC, like history of parental fracture. Where documentation of such variables could not be found in patient charts, the online calculator's default settings were used. Calculated FRAX assessments were subsequently stratified into categories using formulae suggested in Canadian guidelines (i.e., "high risk" = risk of major osteoporotic fracture >20%, "moderate risk" = risk 10%–20%, "low risk" = risk <10%) *(6)*.

Chi-squared tests for independence were used to measure differences in demographics of patients examined by specialists or radiologists. To measure agreement between risk assessments appearing on original reports and algorithmically assigned assessments, percentage agreement (PA) and Krippendorff's alpha (KA) were computed *(24)*. Alpha is similar to the kappa agreement statistic *(25)*, but alpha corrects for kappa's proclivity to positively weigh consistent disagreements *(24)*. Statistics of agreement were computed for all reports and for key subsets to explore variation informed by physician specialty (i.e., radiologist vs specialist), patient age (specifically, for patients <50, 50–65, and >65 yr old), and modifying clinical factors (e.g., history of fracture or steroid use).

Reports associated with disagreements between software and experts were manually reviewed. The lead author (SA) identified and categorized clinical circumstances associated with repeated divergence between expert and computerized assessments; categories were validated by additional members of the research team (RB and AMC).

The Research Ethics Boards of the University Health Network (REB #14-797-CE) and Mount Sinai Hospital (REB #14-0253-C) approved the research study.

Results

Sampled charts contained reports that were read by a total of 6 radiologists and 5 specialists. Table 1 provides demographic information for patients. Specialists saw a significantly larger number of patients on doses of steroids in excess of 7.5 mg of daily prednisone equivalent (3.6% vs 11.4%, p value < 0.01). This discrepancy was more pronounced when comparing numbers of patients on more than a 5 mg prednisone daily equivalent (3.9% vs 23.7%, p value < 0.01). While both specialists and radiologists saw comparable numbers of patients reporting fragility fractures (9.3% for radiologists vs 14.2% for specialists), specialists saw more patients reporting multiple fragility (0.4% vs 5.5%, p value < 0.01) and vertebral fractures (0.0% vs 8.4%, p value < 0.01).

More high-risk patients presented to specialists (4.6% for radiologists vs 14.1% for specialists, p value < 0.05); the same was true for patients under age 50 (12.5% for radiologists vs 26.0% for specialists, p value < 0.05) *(6,9)*. Low-risk patients were more likely to present to radiologists (56.1% vs 21.5%, p value < 0.01). Comparable proportions of patients at "moderate" risk presented to either physician group (21.4% for radiologists, 30.1% for specialists).

Table 2
Fracture Risk Assessments Appearing on Original Reports in the Sample Compared to Algorithmically Assigned Fracture Risk Assessments From the JDMI Software

Expert vs algorithm	High (CAROC)	Moderate (CAROC)	Low (CAROC)	N/A (CAROC)	Total
High	39	6	0	0	45
Moderate	3	122	4	5	134
Low	0	6	195	3	204
N/A	0	0	0	92	92
Total	42	134	199	100	475

Note: The "N/A" category applies to all reports wherein fracture risk was not applicable or could not be calculated due to missing data. Note that 25 reports were omitted from this analysis as they contained expert assessments other than "high," "moderate," or "low" risk.

Abbr: CAROC, Canadian Association of Radiologists and Osteoporosis Canada; JDMI, Joint Department of Medical Imaging.

As Table 1 indicates, several risk categorizations *not* found in guidelines (i.e., neither "high", "moderate" nor "low") were commonly used to describe patients at the margins of moderate risk. A total of 25 reports, or 5.0% of the sample, included such categorizations. Alternative categorizations used by specialists included "moderate to high" risk (on 10 of specialists' reports) or "low to moderate" risk (on 8 reports). Many of these reports (n = 8) were for patients on a 5.0–7.5 mg daily prednisone equivalent dose of steroids; an additional 4 were for patients with *T*-scores on a boundary between risk categorizations. Radiologists occasionally used the category "at least moderate" risk (on n = 7 reports) to describe patients with total hip or spine *T*-scores below –2.5 *(6)*.

Before computing comparative statistics between expert and computer-generated assessments, reports containing categorizations *not* described by Canadian guidelines were excluded. This left a total of 475 reports (203 from specialists and 272 from radiologists), which are presented in Table 2. PA and KA statistics were calculated to relate CAROC and expert assessments on these reports. Agreement statistics were also computed to relate expert and FRAX assessments across the 317 reports containing both "high-," "moderate-," or "low"-risk assessment and the minimum data required for FRAX calculation (i.e., gender, age, height, weight, and femoral neck *T*-scores).

Several statistics of comparison are presented in Table 3. Overall agreement between expert and algorithmically assigned CAROC assessments was high; the PA was 94.3% and the KA indicated excellent agreement (0.92). Most subcategories of reports also were found to have excellent agreement with software calculations. Lower rates of agreement were related to reports for patients with histories of steroid use or fragility fracture; for this group, the PA was 90.2% and the alpha was 0.90. FRAX calculations showed similar patterns of agreement with experts but somewhat lower associated agreement statistics. Across all patients, the PA resulting from these calculations was 80.6% and the alpha 0.63; for those with histories of steroid use or fracture, the PA was 73.6% and the alpha 0.61.

Table 3
Measures of Agreement Between Original and Fracture Risk Assessments Assigned by JDMI Software, by Key Subcategories of Reports

Subcategory of reports	N	Alpha (CAROC)	PA (CAROC)
All reports	475	0.918	94.3
From specialists	203	0.899	92.6
From radiologists	272	0.926	95.6
With risk factors	82	0.860	90.2
No risk factors	391	0.951	96.9
For patients >65 yr old	177	0.886	92.7
For patients 50–65 yr old	185	0.902	95.7

Abbr: CAROC, Canadian Association of Radiologists and Osteoporosis Canada; JDMI, Joint Department of Medical Imaging; PA, percentage agreement.

A total of 27 reports (or 5.4% of the entire sample) contained expert assessments in conflict with suggestions from the JDMI software. Based on manual review, the following categories of discrepancy were established:

1. **Reports containing dictation or transcription errors.** One report contained an error that was the apparent result of automated speech recognition (ASR) software malfunction ("history fragility fracture" was written where "no history of fragility fracture" was likely intended). A second included a history of prednisone at a daily dose of "75 mg" rather than "7.5 mg." Two additional reports contained ASR errors affecting diagnoses. On one, a negative sign was attached to a *T*-score, and on the other, a diagnosis of "no bone mass" was provided.
2. **Reports with ambiguous representation of relevant risk factors.** Examples include one report where "prolonged use of prednisone" was reported without dose information or medication timing. On others, fractures were documented but were not used to calculate risk; it was unclear to abstractors if fractures were true fragility fractures.
3. **Reports for patients with lumbar spine or total hip *T*-scores below –2.5, or with missing femoral neck data.** Three reports contained lumbar *T*-scores below –2.5 yet were without a "moderate" risk assessment as per Canadian guidelines *(3,6)*. Additional reports substituted spine or forearm data for missing femoral neck data in risk calculations. The 2010 guidelines do not provide recommendations as to whether *T*-scores at the spine over –2.5 are related to fracture risk, or how to relate forearm BMD data to risk *(6)*.
4. **Reports for patients at an age or *T*-score boundary.** A total of 6 reports were for patients close to the CAROC's categorical age or *T*-score margins. Of these, 5 were for women <50 (average was 44.8 yr); 1 was for a patient of age 90 (the CAROC does not provide basal risk over age 85). Two additional reports contained *T*-scores close to categorical boundaries (e.g., a *T*-score of –2.3 where –2.3 demarcates "low" from "moderate" risk).

Discussion

Results presented demonstrate that for most routine adult BMD exams, SR software that includes an implementation of the CAROC algorithm suggests fracture risk assessments that are aligned with those of JDMI's experts. However, results also illustrate that experts diverge from assessments suggested by SR about 5% of the time and disagree with them another 5.4% of the time. Much of the variation in experts' reporting of risk in the present study focused on patients in the middle of the risk spectrum, that is, at neither "low" nor "high" risk. Many were on low doses of steroids, were at the age or *T*-score boundaries between CAROC categorizations, or had relatively low measured BMD at the lumbar spine.

Results also show that similar patterns of agreement exist between expert assessments and FRAX calculations; while overall rates of agreement with experts are high, lower rates are associated with patient histories of steroid use or fracture. This finding is consistent with a prior study that validates the CAROC against the FRAX, wherein overall agreement was reported at approximately 89%, whereas agreement for patients with a risk factor was 74% across 1 major Canadian cohort *(8)*. Lower rates of overall concordance reported here may be due in part to JDMI's documentation practices, which are tailored to the CAROC and do not always capture factors of relevance to FRAX calculations (like history of parental fracture). In addition, JMDI's assessment software weighs spine and total hip information, whereas the version of the CAROC that was originally validated against the FRAX did not *(8)*.

Some detected variations in expert assessments promise to be addressable by SR software. These include the following:

1. **Variation due to lack of report clarity.** In the present study, researchers sometimes had difficulty ascertaining the relationship between patients' clinical factors and assessments. A similar difficulty is common among referring physicians *(14,16)*. JDMI's format, as it details assessment "ingredients," promises to limit this kind of difficulty.
2. **Variation due transcription or dictation errors.** In the present study, clinically significant dictation errors were discovered on a few reports from physicians using ASR. In a 2011 study exploring accuracy of ASR and within JDMI, 35% of breast magnetic resonance imaging reports and 13% of interventional radiology reports were found to contain major or minor dictation errors *(26)*. While the error rate reported here is lower, errors were detected only when they *directly* influenced clinical interpretations. JDMI's SR software, as it produces assessments without dictation, promises to limit such errors.

However, some variability proved to be not correctable by SR software as it was related to guideline ambiguity; this must be addressed by clinical consensus. Nevertheless, reports vulnerable to this kind of variability can be recognized and, theoretically, flagged for close review. Examples include reports for patients:

1. **With spine or total hip *T*-scores ≤ –2.5.** The 2010 Canadian guidelines state that low hip or spine *T*-scores place a patient at "at least moderate risk" for fracture *(6)*. Flow diagrams published in the CAR Technical Standards for Bone Mineral Density Reporting suggest risk for these individuals should be "moderate" but no more *(3)*. In addition, clinical practice guidelines mention that risk may be underestimated when spine *T*-scores are "much lower" than the hip *(6)* but do not provide formulae to determine clinically significant discrepancies. Results suggest that

clinicians are sensitive to spine and total hip data but employ varying assessment heuristics in these situations.

2. **With data on a categorical boundary of the CAROC.** Experts in our study often assigned risk assessments to patients who were just under age 50 (where CAROC assessments do not apply). It may be reasonable to do this in some cases; FRAX assessments for 40-yr-olds are similar to those for 50-yr-olds *(7)*. Reports for patients at these boundaries can be readily flagged by software for close review.
3. **On low doses of steroids.** The present study shows that steroid use, particularly at relatively low doses, is associated with variation in "moderate risk" reporting. Research does, in fact, suggest that individuals at low steroid doses are at increased risk of fracture; a 2000 study exploring the relationship between steroids to incidence of fracture found that patients on 2.5–7.5 mg of prednisone daily were >1.7 times more likely to sustain hip or vertebral fractures than patients not on steroids *(27)*. The present study suggests that many clinicians, particularly specialists, consider low doses of steroids but use different assessment heuristics when low doses occur.

Limitations

The sample of reporting physicians in the present study is confined to a few related institutions that perform BMD examinations. Nevertheless, these particular reporting physicians reflect variability across JDMI, which is the second largest provider of BMD examination services in Ontario.

In addition, the understanding of patients' clinical risk factors is based on documentation in charts. The high incidence of patients on low doses of steroids and with multiple fragility fractures presenting to specialists may be an artifact of specialists' documentation practices. Specialists may be more likely to ask patients for details about medication usage and prior fractures, or to examine for vertebral fractures when abnormalities are noted on BMD images.

Finally, many gaps revealed by the present study are not correctable by software as has been explained. Surveys of physicians show acceptance of guideline implementation systems to be diminished when guidelines are poorly justified or ambiguous *(28,29)*. Such ambiguity, however, must be addressed by clinical consensus.

Conclusion

Results demonstrate that for BMD exam reports of adult patients who are not on steroids and have no fracture history, JDMI's SR software produces assessments that reasonably agree with both JDMI experts and FRAX calculations. Moreover, results show that SR software stands to reduce variability in, ameliorate errors in, and improve clarity of BMD reports.

Acknowledgments

The authors thank Atiya Hermaj, Tarik Bereket, Gail Jefferson, Barry Ivo, Shahab Jalali, Leon Goonaratne, Jennifer Caytano, and the Ontario Ministry of Health and Long-Term Care for funding.

References

1. 2013 Official positions of the International Society for Clinical Densitometry. Available at: http://www.iscd.org/official-positions/2013-iscd-official-positions-adult/. Accessed July 30, 2015.
2. Writing Group for the ISCD Position Development Conference. 2004 Indications and reporting for dual-energy X-ray absorptiometry. J Clin Densitom 7(1):37–44.
3. Siminoski K, O'Keeffe M, Lévesque J, et al. 2013 CAR technical standards for bone mineral densitometry reporting. Available at: http://www.car.ca/uploads/standards%20guidelines/20140122_en-bmdreporting.pdf. Accessed July 30, 2015.
4. Siminoski K, Leslie WD, Frame H, et al. 2005 Recommendations for bone mineral density reporting in Canada. Can Assoc Radiol J 56(3):178–188.
5. Siminoski K, Leslie WD, Frame H, et al. 2007 Recommendations for bone mineral density reporting in Canada: a shift to absolute fracture risk assessment. J Clin Densitom 10(2):120–123.
6. Papaioannou A, Morin S, Cheung AM, et al. 2010 2010 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: summary. CMAJ 182(17):1864–1873.
7. Kanis JA. on behalf of The World Health Organisation Scientific Group. 2007 Assessment of Osteoporosis at the Primary Health Care Level. WHO Collaborating Centre for Metabolic Bone Diseases. Available at: http://www.iofbonehealth.org/sites/default/files/WHO_Technical_Report-2007.pdf. Accessed July 30, 2015.
8. Leslie WD, Berger C, Langsetmo L, et al. 2011 Construction and validation of a simplified fracture risk assessment tool for Canadian women and men: results from the CaMos and Manitoba cohorts. Osteoporos Int 22(6):1873–1883.
9. Lentle B, Cheung AM, Hanley DA, et al. 2011 Osteoporosis Canada 2010 guidelines for the assessment of fracture risk. Can Assoc Radiol J 62(4):243–250.
10. Leslie WD, on behalf of The Manitoba Bone Density Program Committee. 2008 Absolute fracture risk reporting in clinical practice: a physician-centered survey. Osteoporos Int 19(4):459–463.
11. Oppermann B, Ayoub W, Newman E, et al. 2010 Consultative DXA reporting improves guideline-driven quality of care-implications for increasing DXA reimbursement. J Clin Densitom 13(3):315–319.
12. Beattie K, Ioannidis G, MacDermid J, et al. 2014 Appropriate osteoporosis treatment by family physicians in response to FRAX vs CAROC reporting: a randomized controlled trial. J Clin Densitom 17(4):458–465.
13. Stock JL, Waud CE, Corderre JA, et al. 1998 Clinical reporting to primary care physicians leads to increased use and understanding of bone densitometry and affects the management of osteoporosis. Ann Intern Med 128:996–999.
14. Allin S, Munce S, Carlin L, et al. 2014 Fracture risk assessment after BMD examination: whose job is it, anyway? Osteoporos Int 25(5):1445–1453.

15. Lewiecki EM, Binkley N, Petak SM. 2006 DXA quality matters. J Clin Densitom 9(4):388–392.
16. Sale J, Bogoch E, Meadows L, et al. 2015 Bone mineral density reporting underestimates fracture risk in Ontario. Health 7:566–571.
17. Allin S, Munce S, Schott AM, et al. 2013 Quality of fracture risk assessment in post-fracture care in Ontario, Canada. Osteoporos Int 24(3):899–905.
18. Schwartz LH, Panicek DM, Berk AR, et al. 2011 Improving communication of diagnostic radiology findings through structured reporting. Radiology 260(1):174–181.
19. Kahn CE Jr, Langlotz CP, Burnside ES, et al. 2009 Toward best practices in radiology reporting. Radiology 252(3):852–856.
20. Taylor A, Hill AN, Binongo JNE, et al. 2007 Evaluation of two diuresis renography decision support systems to determine the need for furosemide in patients with suspected obstruction. AJR Am J Roentgenol 188(5):1395–1402.
21. Burnside ES, Ochsner JE, Fowler K, et al. 2007 Use of microcalcification descriptors in the BI-RADS 4th Edition to stratify the risk of malignancy. Radiology 242:388–395.
22. Stivaros SM, Gledson A, Zeng X, et al. 2010 Decision support systems for clinical radiological practice—towards the next generation. Br J Radiol 83(995):904–914.
23. Canadian FRAX calculator. Available at: http://www.shef.ac.uk/FRAX/tool.aspx. Accessed February 1, 2016.
24. Krippendorff K. 2013 Content analysis: an introduction to its methodology. 3rd ed. Thousand Oaks, CA.
25. Cohen J. 1960 A coefficient of agreement for nominal scales. Educ Psychol Meas 20(1):37–46.
26. Basma S, Lord B, Jacks LM, et al. 2011 Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. AJR Am J Roentgenol 197(4):923–927. doi:10.2214/AJR.11.6691.
27. Van Staa TP, Leufkens HG, Abenhaim L, et al. 2000 Use of oral corticosteroids and risk of fractures. J Bone Miner Res 15(6):993–1000.
28. Teach RL, Shortliffe EH. 1981 An analysis of physician attitudes regarding computer-based clinical consultation systems. Comput Biomed Res 14(6):542–558.
29. Grol R, Grimshaw J. 2003 From best evidence to best practice: effective implementation of change in patients' care. Lancet 362:1225–1230.